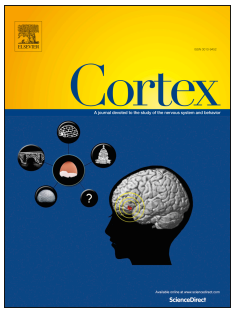
Accepted Manuscript



No evidence for enhanced distractor template representation in early visual cortex

Reshanne R. Reeder, Christian N.L. Olivers, Michael Hanke, Stefan Pollmann

PII: S0010-9452(18)30255-7

DOI: 10.1016/j.cortex.2018.08.005

Reference: CORTEX 2377

To appear in: *Cortex*

Received Date: 27 February 2018

Revised Date: 7 August 2018

Accepted Date: 11 August 2018

Please cite this article as: Reeder RR, Olivers CNL, Hanke M, Pollmann S, No evidence for enhanced distractor template representation in early visual cortex, *CORTEX* (2018), doi: 10.1016/j.cortex.2018.08.005.

No evidence for enhanced distractor template representation in early visual cortex

Reshanne R. Reeder*[1], Christian N. L. Olivers[2], Michael Hanke[3,4], & Stefan Pollmann[1,4]

[1]Department of Experimental Psychology, Institute of Psychology, Otto-von-Guericke University, Magdeburg, Germany

[2]Department of Experimental and Applied Psychology, Vrije Universiteit, Amsterdam, The Netherlands

[3]Psychoinformatics Lab, Institute of Psychology, Otto-von-Guericke University, Magdeburg, Germany

[4]Center for Behavioral Brain Sciences, Otto-von-Guericke University, Magdeburg, Germany

Correspondence to:

Reshanne R. Reeder

Institute of Psychology

Otto-von-Guericke University

Universitaetsplatz 2

39106 Magdeburg

Germany

e-mail: reshanne.reeder@gmail.com

Keywords: negative template, target template, visual search, fMRI, representational similarity analysis

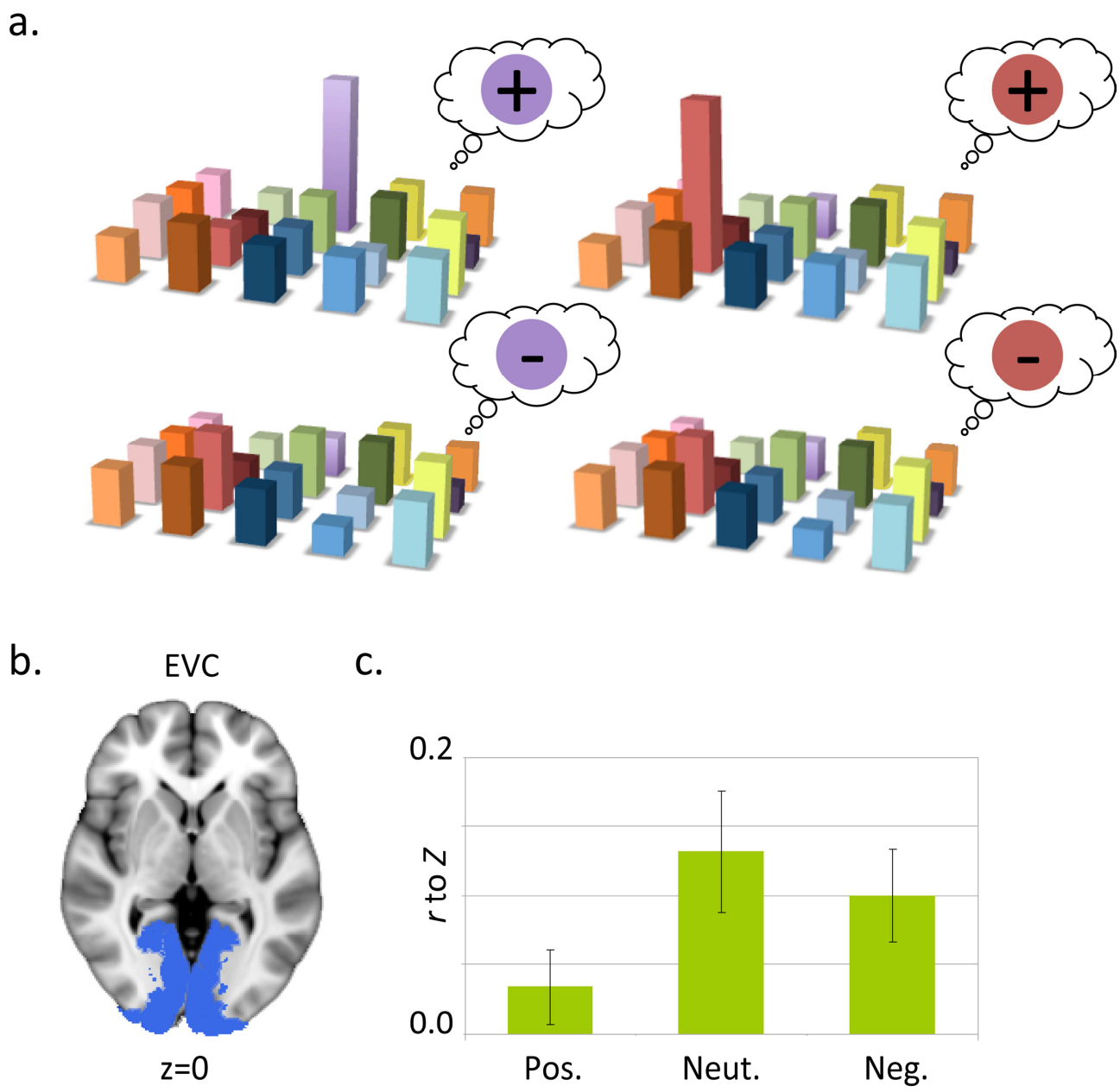
Humans can retain task-relevant visual information in working memory and use it to compare against visual information selected from the environment. Behaviorally, this speeds target detection (Desimone & Duncan, 1995). In the brain, this manifests as distinct target-related cortical activity patterns in visual cortex in preparation for visual stimulation (Gayet et al., 2017; Harrison & Tong, 2009). In recent years, it has been hotly contested whether, in addition to these "target templates", the brain also makes use of "templates for rejection" – representations of distracting information to benefit subsequent target detection (Arita et al., 2012; Beck & Hollingworth, 2015; Beck et al., 2017; Moher & Egeth, 2012; Reeder et al., 2017). A difference in how distractor information is retained compared to target information would suggest the use of a different preparatory template. Our questions for the current paper focus on the nature of such preparatory representations of targets and distractors for search. Is a distractor represented distinctly like a target in visual working memory (VWM) with an additional "tag" that this should be rejected once it has been identified during search? Or is there no distinct representation, perhaps even suppression, of the distractor feature in sensory brain areas during the preparatory period? The current study is the first to provide evidence that visual features of anticipated distractors are not represented more distinctly than irrelevant features (that will not appear in the search display) in early visual cortex (EVC), supporting the hypothesis that a template for rejection is functionally different from a target template.

We asked subjects to detect a target in an array containing four items of one color and four items of a second color, while undergoing fMRI (see Figure S1 in the Supplementary Methods). One of the two colors was cued beforehand as positive ("the target will appear in this color"), negative ("only distractors will appear in this color"), or neutral ("this color will not appear in the search display"). The two colors that appeared in the search display on each trial were chosen from a selection of five colors. Each of the five colors appeared as a positive, negative, or neutral cue an equal number of times. Initial univariate analyses of the blood-oxygenation-level-dependent (BOLD) signal revealed a lower BOLD response for negative cues compared to positive and neutral cues in EVC, despite a behavioral benefit

to having foreknowledge of the upcoming distractor color (Reeder et al., 2017). This pattern is not predicted by the equal representation plus negative tag hypothesis, but rather supports the distractor inhibition hypothesis of templates for rejection. Nevertheless, the regional activation modulation that we reported previously is too unspecific to demonstrate differences in target and distractor feature representations. For instance, a stronger BOLD response following positive cues may reflect a global preparatory increase of neuronal activation instead of a selective increase of activation in those neurons representing the target feature. Likewise, a drop in the regional BOLD amplitude may reflect inhibition of preparatory attention rather than selective suppression of the cued feature. If these regional BOLD amplitude modulations are driven by feature-selective modulation of neuronal delay activity, we should see more distinct patterns of activity for target templates than for task-irrelevant features. If distractor templates are characterized by inhibition of the distractor feature rather than facilitation, they should lack the distinctiveness of target templates. Instead, distractor features should show comparable distinctiveness to irrelevant features (Figure 1) or even an anti-correlated pattern, i.e., decreased activation in voxels that show increased activation from baseline when the same feature is cued as a target.

Figure 1. a.) An illustration of the study's hypotheses: the activation of a target template (represented by "+" inside the colored bubble) leads to selective activation of EVC neurons representing the target color. This, in turn, leads to distinct activation patterns for the different target colors. Contrarily, the activation of a template for rejection (represented by "—" inside the colored bubble) leads to decreased activation of EVC neurons and therefore decreased variability in stimulus-related activity. Thus, different negatively cued colors will elicit weaker activity patterns that are more similar to those elicited by task-irrelevant colors. b.) A brain in MNI space showing the extent of the EVC region analyzed (in blue). Left hemisphere is displayed on the right. c.) A bar graph showing the average *r* to *Z* values across colors presented as positive, neutral, and negative cues. Error bars represent the

standard error of the mean.



To investigate how distinctly the cue colors were represented in preparation for search, we used representational similarity analysis (RSA; Kriegeskorte et al., 2008) combined with a searchlight method implemented in PyMVPA (Hanke et al., 2009) within EVC (see Supplementary Methods for a detailed description of the analysis pipeline). Correlation distance (1-*r*) between beta weights was calculated for 15 conditions of interest (3 cue types x 5 colors), serving as the "distinctiveness" metric.

The data expressing the non-transformed 1-*r* values associated with each color and cue type, as well as the mean 1-*r* value for each cue type collapsed across colors, are reported in Table 1. 1-*r* values were then transformed into *r* values and Fisher *Z*-transformed to ensure a normal distribution of the data for statistical hypothesis tests. Lower *Z* values therefore indicate smaller correlations (i.e., greater distinctiveness) between color representations (see Figure S2).

We first performed a 3 (cue type: positive, negative, neutral) x 5 (color) repeated-measures ANOVA to test for representational distinctiveness of the different colors for each cue type. This revealed a significant main effect of cue type ($F(2,32)=4.960$, $p=0.013$, $\eta^2_p=0.237$), no main effect of color ($F(2,32)=1.237$, $p=0.304$, $\eta^2_p=0.072$), and no interaction between the two ($F(2,32)=0.812$, $p=0.593$, $\eta^2_p=0.048$). We then collapsed the data across color and conducted paired-samples t-tests to gauge the representational distinctiveness differences between cue types.

Table 1

Mean 1-r *distinctiveness values for each color and their standard deviation (SD). The mean and SD of each cue type with all colors combined are shown in the last column*

	Light Pink		Orange		Chartreuse		Cyan		Orchid		All colors	
Cue type	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Positive	1.004	0.212	0.989	0.109	1.018	0.115	0.949	0.153	0.872	0.143	0.966	0.109
Negative	0.897	0.154	0.901	0.123	0.885	0.192	0.895	0.164	0.931	0.227	0.902	0.133
Neutral	0.828	0.190	0.853	0.177	0.892	0.228	0.898	0.214	0.903	0.207	0.875	0.167

Note. The SD for all colors combined was calculated as the between-subjects SD after obtaining the mean 1-r *values collapsed across the 5 colors. Hex codes of each color are provided in the Supplementary Methods.*

To test for increased distinctiveness of cued target features, we first conducted a paired-samples t-test on positive > neutral cue distinctiveness in EVC. Positive cues were represented more distinctly than neutral cues ($t(16)=-2.574$, $p=0.01$, one-tailed, $d=0.953$; Cohen's *d* is corrected for dependent samples; Morris & DeShon, 2002). We then conducted a paired-samples t-test on positive > negative cue distinctiveness, which also showed a significant difference between cue types ($t(16)=-2.180$, $p=0.023$, one-tailed, $d=0.813$). We then tested if cued distractor features were more distinctively

represented than neutral features. A paired-samples t-test showed no difference between negative and neutral cue distinctiveness ($t(16)=-0.794$, $p=0.216$, one-tailed, $d=-0.198$). We followed this up with a Bayes factor (BF) analysis[1] (JASP Team, 2018). Bayes factors are particularly useful to test if the lack of a significant difference is due to equal distinctiveness or low power (see Dienes, 2014). With the Cauchy prior set to the default of 0.707, we found a BF01 = 3.04 (in favor of the null hypothesis), which provides moderate support for equal distinctiveness of negative and neutral cues (Schönbrodt & Wagenmakers, 2018). A BF robustness test on these data revealed that the likelihood of this lack of a difference increases when the Cauchy prior is increased, suggesting this effect survives variability in the prior width.

These analyses looked at the distinctiveness of the five colors given a cue condition, but we can also look at the similarity of activation patterns across cue conditions for a given color. For example, a positive cue may facilitate firing in a neuron that codes a given color and inhibit firing in a neuron that codes a different color, whereas a negative cue may inhibit firing below baseline in the former and increase firing in the latter. This could lead to comparable distinctiveness in the above pattern analyses between colors, but a negative correlation for the same color across cue conditions. We therefore tested the correlation between positive and negative cue distinctiveness and found a moderate positive correlation ($r=0.51$, $p=0.036$), supporting the hypothesis that negative cue activation patterns are less distinct, but qualitatively similar to positive cue representations, rather than inverted (which would be suggested by a negative correlation).

These results show that only preparatory target feature representations in EVC are more distinct than neutral feature representations, whereas the preparatory representation of distractor features are not distinct from neutral feature representations. This pattern is incompatible with the hypothesis that target and distractor representations are both enhanced by attention during the preparatory period. The

[1] For completeness, we report the other comparisons: positive vs. neutral cues BF01 = 0.33, and positive vs. negative cues BF01 = 0.62.

current results suggest that the lower univariate BOLD signal for negatively cued colors compared to positively cued colors in the previous analysis of this dataset (Reeder et al., 2017) was driven by a general suppression of visual processing in EVC, rather than color-specific suppression.

While distractor colors and irrelevant colors showed similar levels of distinctiveness, we observed no negative correlation between positively and negatively cued colors that would have resulted if neuronal activation of the distractor feature was a mirror image of the pattern elicited by target features, i.e., distractor suppression in neurons where there is target facilitation, and vice versa. We think this did not occur because of the overall low level of activation during the delay between cue offset and search onset. While attention can cause clear increases and decreases during sensory stimulation (e.g. Treue & Martinez Trujillo, 1999), during the delay period (which we have analyzed here), neuronal activity is typically much reduced even if the cue matches the preferred feature of the neuron (Bichot et al., 2005; Chelazzi et al., 1993; see also decoding of working memory content in the absence of an elevated BOLD-response: Harrison & Tong, 2009; Serences et al., 2009). Inhibition of stimulus features would only reduce the neuronal firing rate from already low delay activity to zero, leaving much less room for distinctive differences in firing rate compared to the facilitatory modulation of firing rate by positive cues.

At this point we cannot rule out that features cued as distractors may be represented by a negatively correlated pattern with target features, but our current methods lack the sensitivity to measure it. Increasing sensitivity, e.g., by using higher magnetic field strength fMRI, may lead to further insights. Moreover, note that our irrelevant feature baseline may itself represent inhibition of EVC. The facilitation of reaction times by negative cues compared to neutral cues yields no indication that inhibition was only present in the former. It may simply be due to the fact that distractor inhibition is useful for search whereas inhibition of irrelevant features is not. Thus, defining a different neutral baseline that is less likely to induce inhibition may also be a way to address the effects of distractor inhibition on EVC representations. Finally, it would be worthwhile to investigate whether the

modulation of representational distinctiveness depends on the features that are used as cues.

The present results show that target templates were distinctly represented in EVC, whereas no distinct representation was observed for distractor templates, compared to baseline. Combined with the previous results showing region-wide preparatory target facilitation and distractor inhibition, we conclude that the representation of target templates and templates for rejection reflect differences in both global and feature-selective brain activity.

References

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 580–584.

Beck, V. M., & Hollingworth, A. (2015). Evidence for negative feature guidance in visual search is explained by spatial recoding. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(5), 1190.

Beck, V. M., Luck, S. J., & Hollingworth, A. (2017). Whatever You Do, Don't Look at the...: Evaluating Guidance by an Exclusionary Attentional Template. *Journal of experimental psychology. Human perception and performance*.

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529-534.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345-347.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193-222.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, *5*, 781.

Gayet, S., Guggenmos, M., Christophel, T. B., Haynes, J. D., Paffen, C. L., Van der Stigchel, S., & Sterzer, P. (2017). Visual working memory enhances the neural response to matching visual input. *Journal of Neuroscience, 37*(28), 6638-6647.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: a python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics, 7*(1), 37-53.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature, 458*(7238), 632-635.

JASP Team (2018). JASP (Version 0.8.6)[Computer software].

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis–connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*.

Moher, J., & Egeth, H. E. (2012). The ignoring paradox: Cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. *Attention, Perception, & Psychophysics, 74*(8), 1590-1605.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological methods, 7*(1), 105-125.

Reeder, R. R., Olivers, C. N., & Pollmann, S. (2017). Cortical evidence for negative search templates. *Visual Cognition, 25*(1-3).

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological science, 20*(2), 207-214.

Treue, S., & Martinez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399*(6736), 575-579.

Author Contributions

RRR, CNLO, and SP conceived the experiment and wrote the paper. RRR designed and conducted the

experiment and performed the analyses. MH provided input on all analyses and RRR and MH wrote the Supplementary Methods.

Acknowledgements

We would like to thank Emanuele Porcu for his help with the fMRI preprocessing and Lasse Güldener for his help with data collection. This project was supported by Open Research Area grants DFG PO 548/16-1 to SP and NWO 464-13-003, NL, and European Research Council Consolidator grant ERC-CoG-2013-615423 to CNLO.

Competing interests

The authors declare no competing interests.